OHIO RIVER BASIN PRECIPITATION FREQUENCY STUDY
Update of *Technical Paper 40*

Fourth Progress Report
for the Period
April 1, 1999 through February 29, 2000

Lesley Tarleton Julian
Julie M. Daniel, Bingzhang Lin, David Riley and Michael Yekta

Office of Hydrology
U.S. National Weather Service
Silver Spring, Maryland

March 2000

DISCLAIMER

The data and information presented in this report should be considered as preliminary and are provided only to demonstrate current progress on the various technical tasks associated with this project. Values presented herein are NOT intended for any other use beyond the scope of this progress report. Anyone using any data or information presented in this report for any purpose other than for what it was intended does so at their own risk.

TABLE OF CONTENTS

OHIO RIVER PRECIPITATION FREQUENCY STUDY
Update of *Technical Paper 40*

Fourth Progress Report
for the Period
April 1, 1999 through February 29, 2000

I. STUDY OVERVIEW

A. Purpose and Scope

The Hydrometeorological Design Studies Center, Office of Hydrology, U.S. National Weather Service is performing a precipitation frequency study to update Technical Paper No. 40, *Rainfall Frequency Atlas of the United States for Durations from 30 Minutes to 24 Hours and Return Periods from 1 to 100 Years* for the Ohio River basin. The study involves the completion of certain specific tasks including collecting and performing quality control of data, compiling and formatting datasets for analyses, selecting applicable frequency distributions and fitting techniques, analyzing data, mapping and preparing reports and other documentation.

B. Study Area

The study area covers 13 states completely and parts of nine additional bordering states. The Susquehanna River and Delaware River basins are also included in the study area.

Currently, the study area is divided into 16 near-homogeneous climatic (i.e., defined as extreme precipitation climate) regions. Factors considered in defining the regions include 1) the season (or seasons) of highest precipitation, 2) the type of precipitation (e.g., general storm, convective, tropical storms or hurricanes, or a combination), 3) the climate, 4) the topography (especially as it interacts with the weather systems) and 5) the homogeneity of these factors in a single area. The regions may be redefined during the course of the study.

The study area is displayed in Figure 1. The core and border states and regional boundaries are shown on the figure.

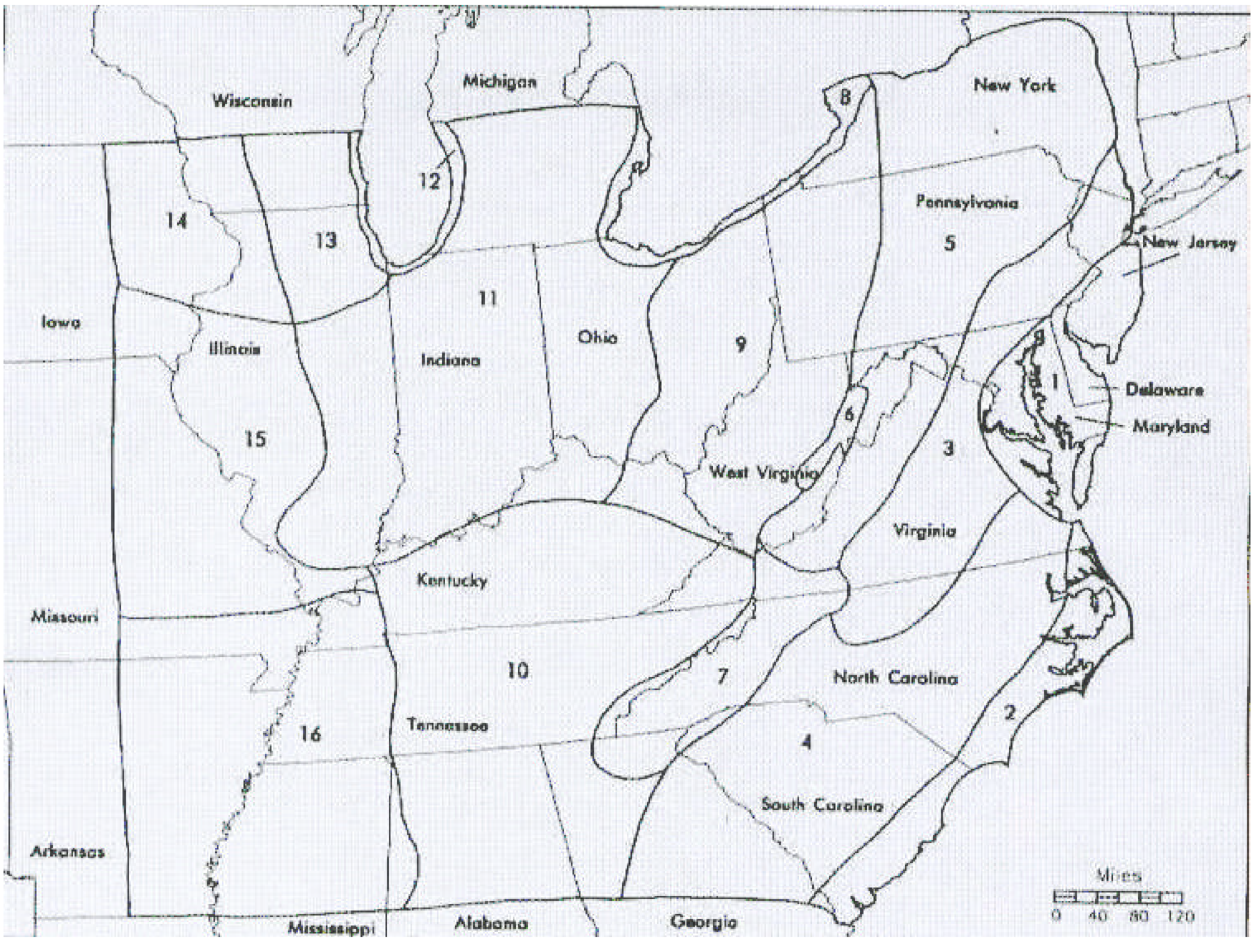

Figure 1. Ohio River study area and preliminary region boundaries.

II. TASK STATUS

The following sections discuss the status of each ongoing task and provide a short technical description of decisions made and task accomplishments, when applicable. Project tasks not discussed are in the planning stages or have yet to start.

A. Data Collection and Quality Control

The datasets are nearly complete. The compilation includes quality control, merging stations where appropriate and formatting for analysis. For merging, stations must be: 1) within 1000 feet in elevation, 2) within 0.070 latitude and longitude in distance, 3) contain a gap between records of 60 months (5 years) or less. The following sections discuss the status of ongoing or completed activities involving daily, hourly, 15-minute and n-minute datasets.

1. Daily Data

Daily data from the National Climatic Data Center (NCDC), the U.S. Army Corps of Engineers (COE) and the U.S. Geological Survey (USGS) were updated through November 1998. The updating includes quality control, elimination of duplicate records and merging stations for longer station records where appropriate.

Table 1 shows the number of NCDC stations merged in each state. The merging can extend station records and/or eliminate redundancy. A total of 278 stations have merged records. Core states are shown in bold; the others are border states.

Table 1. Number of NCDC daily stations with merged data.

STATE	NO.
Alabama	6
Arkansas	5
Delaware	0
Georgia	2
Iowa	1
Illinois	23
Indiana	9
Kentucky	25
Maryland	15
Michigan	4
Missouri	4
Mississippi	3
North Carolina	21
New Jersey	10
New York	27
Ohio	22
Pennsylvania	36
South Carolina	6
Tennessee	16
Virginia	20
Wisconsin	3
West Virginia	20
TOTAL	278

Table 2 has the complete list of daily stations from all sources: NCDC, COE and USGS. Only stations with 20 years or more will be used for frequency analysis. The average years are based on all stations. The total station column includes all stations, including those with short records. Figure 2a shows all the daily and hourly stations, including those with short records. Figure 2b shows daily stations; Figure 2c shows hourly stations.

Table 2. Daily dataset for state stations using source data (core states in bold).

STATE ID	TOTAL STATIONS	STATIONS $\geq$ 20 YEARS	AVERAGE YEARS
Alabama	160	107	35
Arkansas	128	88	43
Delaware	19	10	35
Georgia	145	111	43
Iowa	149	86	40
Illinois	381	254	47
Indiana	326	195	42
Kentucky	382	229	37
Maryland	164	89	34
Michigan	143	78	40
Missouri	196	151	45
Mississippi	111	86	46
North Carolina	367	266	40
New Jersey	120	89	46
New York	361	218	33
Ohio	388	260	45
Pennsylvania	525	319	34
South Carolina	152	115	46
Tennessee	510	260	26
Virginia	317	213	35
Wisconsin	86	59	47
West Virginia	292	153	30
TOTALS	**5422**	**3436**	

2. Hourly Data

- The hourly data obtained from the COE have been reformatted and appended to the hourly stations for the periods with a gap in the dataset.

- Hourly data through November 1998 are included in the dataset. Some of these later data must still be added and/or merged before the datasets are complete and ready for analysis.

- Some COE hourly stations are co-located with daily stations. If a co-located hourly station had a longer record than the daily, the hourly values were recomputed as 24-hour (daily) totals and the additional years appended to the daily station record. NCDC daily records were also extended in the same method.

Table 3. Number of hourly stations merged using NCDC data (core states in bold).

STATE ID	NO.
Alabama	4
Arkansas	4
Delaware	0
Georgia	4
Iowa	5
Illinois	22
Indiana	22
Kentucky	18
Maryland	1
Michigan	3
Missouri	7
Mississippi	3
North Carolina	12
New Jersey	4
New York	12
Ohio	31
Pennsylvania	56
South Carolina	2
Tennessee	9
Virginia	15
Wisconsin	3
West Virginia	9
TOTAL	246

- The breakdown of the data received from the COE Districts follows:

  Huntington (78 stations):
  Data from 68 stations were appended to the NCDC data. Ten stations had too short (<20 years) records and were not co-located. The short-record stations may be useful for storm analysis.

  Nashville (101 stations):
  Data from 69 stations were appended to the NCDC data. Thirty-two stations had too short (<20 years) records and were not co-located. The short-record stations may be useful for storm analysis.

Louisville (20 stations):
Data from two stations were appended to the NCDC data. Thirteen stations had too short (<20 years) records and were not co-located. Data for five stations were not used since the NCDC had already provided these data. The short-record stations may be useful for storm analysis.

Table 4. Hourly dataset for state stations using source data (core states in bold).

STATE ID	TOTAL STATIONS	STATIONS $\geq$ 20 YEARS	AVERAGE YEARS
Alabama	31	20	31
Arkansas	34	26	34
Delaware	4	3	32
Georgia	67	35	27
Iowa	44	26	28
Illinois	141	84	28
Indiana	127	76	28
Kentucky	123	63	25
Maryland	38	19	21
Michigan	44	25	28
Missouri	74	53	33
Mississippi	44	27	32
North Carolina	106	54	25
New Jersey	49	26	25
New York	91	54	27
Ohio	175	110	29
Pennsylvania	236	151	28
South Carolina	42	27	33
Tennessee	137	50	20
Virginia	91	54	26
Wisconsin	22	16	32
West Virginia	81	47	27
TOTALS	**1801**	**1046**	

3. 15-minute Data

- The 15-minute dataset from the NCDC is unusable, as it appears to be mainly hourly data in a 15-minute format. The following is a sample of 15-minute data for January 16, 1991 for station 31-1458, Cape Hatteras, North Carolina.

15-minute data example:

31-14581991 116 0 0 0 0 0 0 0 0 ***23 0 0 0 74*** 0 0 0 2 0 0 0 65 0 0 0 15 0 0 0 25 0 0 0 27 0 0 0109 0 0 0195 0 0 0 25 0 0 0 23 0 0 0 12 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.

The numbers in Italic print, 23 and 74, correspond to the 15-minute data. The three zeros in between the numbers represent place holders for the 15-minute steps. Thus, it simply represents hourly data. This is a consistent pattern throughout the data from this station, and roughly 90 percent of the other stations in the dataset have data of this format. The following is a sample of hourly data for the same station.

Hourly data example:

31-14581991 116 0 ***23 74*** 2 65 15 25 27 109 195 25 23 12 0 0 0 0 0 0 0 0 0 0 0.

- Additionally, numerous values could not have occurred in 15 minutes, such as 8.00 inches in midwinter in northern Indiana. Contact has been made with the data source, the NCDC, about these concerns. The NCDC verified that the 15-minute dataset obtained is their best and most up-to-date.

- Roughly 100 of 668 stations with 15-minute data were thoroughly examined. Of these stations approximately 10 contained actual 15-minute data. Since so few true data from 15-minute dataset exist, an analysis of these data would not yield useful results.

4. N-minute Data

The n-minute data are ready for further analysis. Digital n-minute data for 76 stations have been obtained from NCDC in two different datasets, which have been merged into a common format. Previously, software had been written to compute annual maximum and partial duration series for 5-, 10-, 15-, 30-, 60-, and 120-minute durations, which will be input to the L-moment program. The n-minute data will also be used to compute the conversion factor for 1 hour to 60 minutes.

B. Frequency Distribution Fitting Analyses

This task is to evaluate and select the frequency distribution(s) which provides the best fit for the data. Based on the work of Hosking and Wallis (1997), earlier papers by the same authors, and evaluation by the Hydrometeorological Branch (predecessor of HDSC), Linear-moments (L-moment) were determined to provide the best method of developing precipitation frequency estimates. For this project, based on L-moment analyses by Bingzhang Lin (see appendix), it appears that the best fit is the Generalized Normal (GNO) for precipitation frequency computations for the Ohio River basin. IMPORTANT NOTE: In the appendix the Lognormal (LNO) should be Generalized Normal (GNO).

C. Precipitation Frequency Value Calculations

The purpose of this task is to obtain a consistent set of precipitation-frequency values and relations. Part of the procedure includes defining near-homogeneous regions. Initially, sixteen regions have been defined and are shown in Figure 1. As the Ohio River basin has fairly similar conditions over large areas, it may be possible to decrease the number of regions, and thus simplify the analysis. Preliminary calculations have been made using the existing regions. However, testing will continue to determine the final regions for analysis. In that light, the following is a discussion of some of the aspects of L-moment procedures, and also their application to regionalization.

L-moment analysis.

L-moment statistics are used for quality-control and return frequency estimates (Hosking and Wallis 1997). A part of the L-moment analysis is to determine the homogeneity of the areas, and to make any necessary adjustments to improve their homogeneity. L-moment definitions and tests for Discordancy (D), Heterogeneity (H), and Goodness-of-fit are described here.

Discordancy.

Initially, the discordancy measure is used for data checking and quality control. However, in evaluating regions, it is used to determine if a site has been assigned to the appropriate region. It is based on L-moments (L-coefficient-of-variation (L-CV), L-skewness (L-SK), and L-kurtosis (L-KT)), which represent a point in 3-dimensional space, for each site. Then, discordancy (D) is a function of the distance from the cluster of points for the sites in the region being tested. The cluster center is in fact the unweighted mean of the three moments for the sites within the region being tested. Sites with a discordancy value of 3 or greater are considered discordant, and should be examined to see if they possibly belong in another region or have a data problem. The threshold value of 3 is not a rigorous test, but a reasonable level to be expected within a homogeneous region.

Heterogeneity.

Actually, the heterogeneity test consists of three parts, one (H-1) based on L-CV, the second (H-2) based on L-CV and L-SK, and the third (H-3) based on L-SK and L-KT. As in the discordancy test, there is also a threshold value; Hosking and Wallis (1991) recommend a threshold of 1. However, they used wind data in establishing this threshold, and later conversations with Wallis (personal communication 1993) indicate that a threshold of 2 is reasonable, especially for precipitation data. Therefore, for each H-test, a value greater than 2 indicates heterogeneity (H>2), rather than homogeneity (H<2). In general, H-1, based on L-coefficient of variation (L-CV) is most stringent. As precipitation data are highly variable in any case, the heterogeneity results were considered giving less weight to the L-CV criterion.

Goodness-of-fit.

This test measures the "distance" of L-moment statistical parameters of a dataset from various theoretical probability distributions. The threshold for goodness-of-fit tests is 1.64 (absolute value), and 'best-fit' values are those less than or equal to the threshold.

Regionalization.

The initial regionalization may need some refinement. The next steps are being taken and include:

- Check all discordant stations and adjust regions as necessary.
- Review boundaries of heterogeneous regions and adjust or divide as appropriate.

- Run L-moment tests on revised regions and evaluate results.

- Compute return frequencies and compare with theoretical distributions with a real-data-check, Lin and Vogel (1993).

The precipitation frequency analysis for the Update of TP40 project will be based on the best regionalization possible.

Precipitation Frequency Values.

Some sample stations were used to compare Update results with TP40. Table 5 shows the comparison of the Update with TP40 for 6-hour data. For comparison, the 100-year data are italicized.

Table 5. Update and TP40 comparison of 6-hour data.

							New - 6-hr		TP40 - 6-hr	
State	ID	Lat.	Lon.	El. (ft)	Por.		2-yr	*100-yr*	2-yr	*100-yr*
Illinois	11-5136	41.57	88.08	590	20		2.185	*5.300*	2.18	*4.34*
Illinois	11-1664	38.52	88.40	450	46		2.372	*5.753*	2.50	*4.83*
Indiana	12-5535	41.17	86.90	690	46		2.211	*5.149*	2.14	*4.23*
Indiana	12-6697	38.40	86.12	770	49		2.179	*5.076*	2.30	*4.47*
Ohio	33-4403	39.72	82.60	840	47		1.882	*4.585*	1.92	*3.90*
West Virginia	46-8614	38.23	80.88	1760	37		1.674	*3.845*	2.10	*4.50*
West Virginia	46-8777	39.45	79.55	2630	49		1.717	*3.942*	2.00	*4.40*
Kentucky	15-8482	37.42	82.80	890	22		1.614	*3.655*	2.21	*4.50*
Kentucky	15-6580	37.12	87.87	500	22		2.446	*5.538*	2.53	*4.89*
Tennessee	40-4950	35.82	83.98	880	51		1.978	*4.329*	2.55	*4.97*
Tennessee	40-1663	36.32	87.22	390	37		2.472	*5.411*	2.52	*4.90*

Table 6 is a comparison of the Update with TP40 - 24-hour data. For comparison, the 100-year data is italicized.

Table 6. Update and TP40 comparison of 24-hour data.

								New - 24-hr		TP40 - 24-hr	
State	Rgn.	ID	Lat.	Lon.	El. (ft)	Por.		2-yr	*100-yr*	2-yr	*100-yr*
Illinois	11	11-8684	39.80	88.28	650	102		3.040	*6.869*	3.10	*6.3*
Indiana	11	12-5174	41.28	87.42	670	36		3.305	*7.467*	2.80	*5.8*
Indiana	11	12-6056	39.20	86.25	620	36		3.263	*7.373*	3.07	*6.0*
Michigan	11	20-0230	42.30	83.72	900	119		2.227	*5.031*	2.43	*4.8*
Ohio	11	33-8552	40.10	83.78	1000	99		2.632	*5.948*	2.72	*5.2*
Kentucky	11	15-3203	38.42	84.88	490	75		3.133	*7.080*	3.11	*6.1*
Ohio	9	33-3780	41.30	81.15	1230	113		2.339	*5.212*	2.30	*4.4*
Ohio	9	33-0279	39.33	82.10	700	58		2.502	*5.574*	2.52	*4.9*
Pennsylvania	9	36-9318	40.18	80.18	1300	44		2.312	*5.151*	2.57	*5.2*
West Virginia	9	46-2462	37.87	81.47	1260	50		2.372	*5.285*	2.75	*5.6*
Kentucky	10	15-7049	36.85	86.88	570	75		3.614	*7.736*	3.40	*6.7*
Tennessee	10	40-6104	35.63	84.02	960	40		3.168	*6.781*	3.40	*6.8*
Tennessee	10	40-4223	35.55	87.55	980	61		4.313	*9.233*	3.65	*7.1*

III. INTERNET-BASED GRAPHICAL USER INTERFACE

Work on an Internet-based graphical user interface (GUI) for analyzing point precipitation frequency data is nearing completion. Designed after the on-line Alabama rainfall atlas (http://bama.ua.edu/~rain/), the development of the Hydrometeorological Design Studies Center (HDSC) GUI is taking place at the University of Alabama. The Java-coded GUI will first be implemented this spring for the Semiarid Southwest study. However, the GUI has been designed to accommodate the Ohio River Precipitation Frequency Study.

- Using a point-and-click interface, the user chooses a point of interest from a shaded relief map complete with roads, cities and political boundaries. The user also selects duration (short [5min to 24 hours] or long [24 hours to 10 days]), units (inches or millimeters) and season (warm, cool or all). Based on these selections and the latitude/longitude pair, a color intensity-duration-frequency (IDF) curve (hyetograph) and data table are generated. Both the table and graph are printable from a web browser, while the data are downloadable as text for further analysis.

- The underlying data from which the GUI operates are from an ASCII raster dataset, essentially a grid of regularly spaced precipitation frequency estimates. The grid spacing is 30-arc seconds, or 0.00833 degrees. Web links will be placed on the GUI for users wanting to download the basic spatial Geographical Information Systems (GIS) datasets.

- Once implemented, the GUI will be maintained at the National Weather Service Headquarters. The GUI will run on a new, dedicated HDSC web server to provide users with the highest speed and reliability possible. New datasets will be added as they become available.

IV. INTERAGENCY MEETING SYNOPSIS
29 June 1999

As a result of various questions about the Third Progress Report (TPR) for the *Ohio River Basin Precipitation Frequency Study, Update of Technical Paper 40* (February 1999), the Hydrometeorological Design Studies Center (HDSC) met with a Federal interagency committee to discuss the state of the project and consider questions or problems. Agencies represented included the COE, TVA, NRCS, FEMA and FERC. Participants are included in the Appendix.

- Re-TP40. Concern was raised that values in the examples in the TPR were much lower than TP40, particularly at the 100-year level. Thus, it was proposed that the TP40 values be redone or reconstructed (Re-TP40) with the original (TP40) data and methodology to determine the effects of additional data and/or changes in methodology. Insofar as possible, the TP40 values were reconstructed, using the same time period and the Gumbel distribution.

- Re-TP40. TP40 (1961) used the full record through 1958 for daily data, and 1938 through 1957 for hourly data. As not all the TP40 data are digitized, the same records could not be used. However, all available digitized data were used, including hand-entered data up through 1958 (daily) and 1957 (hourly). TP40 used Annual Maximum (AM) series and the Gumbel (GUM) distribution. The results were then converted to partial duration using the reciprocal of the factors given on page 3 in the original TP40. The results were 1.136 for 2-yr, 1.042 for 5-yr and 1.01 for 10-yr. Thus, the maps in TP40 (1961) are partial duration values. The same (AM series, GUM distribution) process was applied to the Re-TP40, with the exception that L-moments were used for the estimation of distribution parameters, whereas TP40 used Conventional Moments.

- Update. The Update of TP40 uses the full data record for all stations through 1998. A partial duration (PD) series is prepared for each station and subject to analysis using L-moment statistics. Daily, 24-hour and 6-hour data series are used. The L-moments permit regional analysis; the results are still point values for individual stations. The regions used in this case are the states of Ohio, Indiana and Illinois and one combined region including all these three states. For the final study, regions will be determined based on their near-homogeneity with regard to extreme precipitation climatology and topography as well.

- Update. After converting daily data to 24 hours, the converted daily and the recorded 24-hour data were combined as one dataset (PD series). Various tests, namely L-moments/Xtest, graphical test and real-data-check, were performed for goodness-of-fit for each of the three states and the combined region. As a result, the Generalized-normal (GNO) was the best fit for the

Update. In TPR the Generalized Pareto Distribution (GPA) was used. Return frequency values were determined using the three best-fit distributions and compared with the original TP40 (1961) for the same stations used in TPR.

- Update. For the 6-hour analysis, the same procedure was followed, but without the need to adjust the 6-hour data to observation day.

- In both the Update and Re-TP40, for the 24-hour analysis the daily data were adjusted to 24 hours, using the same conversion factor, specifically 1.13, as developed in TP40. This compensates for the unfortunate habit of the highest 24-hour amounts occurring over different 24-hour periods, not necessarily the standard observation period. For the future report, a specific study on 1-day versus 24-hour rainfall data for the Ohio River Basin is needed.

- In both the Update and Re-TP40, for the plot-position formula TP40 used $P_m = m/(n+1)$ while the Re-TP40 used $P_m = (m+A)/(n+B)$ with parameters $A = -0.35$ and $B = 0$. Here, m is the m^{th} smallest and n is the size of data series. The Update of TP40 used the same plot-position formula with the same parameter values as for the Re-TP40.

Table 7 shows a comparison of the various approaches to precipitation frequency for the Ohio River Basin Study.

Table 7. Comparison of data and methodology used in Four Parallel Approaches for Ohio Project.

	Four parallel approaches			
	TP40	Third Progress Report	Re-TP40 (1999)	Update-TP40
Data length	(......) – 1958	(......) – 1996	(......) – 1958	(......) – 1998
Data type	Annual Max.	Annual Max.	Annual Max.	Partial Duration
Method of estimation of distribution parameters	Conventional Moments	L-Moment	L-Moment	L-Moment
Plot-Position Parameters *	A = 0, B = 1	A = 0, B = 0	A = -0.35, B = 0	A = -0.35, B = 0
Conversion factor of Daily-to-24hour	1.13	None	1.13	1.13
AM-to-PD **	Yes	No	Yes	N/A
Goodness-of-fit	(Unclear)	3 methods used	3 methods used	3 methods used
Best distribution	Gumbel	GLO	GPA	GNO ***

Notes:

Re-TP40 – Reconstruction of TP40 keeping the same data, while using L-Moment procedures.

* Plot-Position formula used here is $P_m = (m+A) / (n+B)$ in which m is the m^{th} smallest and n is the size of the data series.

** The factors used to convert Annual Maximum to equivalent Partial Duration are: 1.13 for 2-year, 1.04 for 5-year and 1.01 for 10-year. No conversion for return frequencies higher than 10-year.

*** GNO – Generalized Normal distribution which is the same as LNO presented in the Interagency Meeting for Update of TP40/Ohio River Basin, 29 June 1999, in Silver Spring.

GPA – Generalized Pareto distribution
GLO – Generalized Logistic distribution
AM – Annual Maximum
PD – Partial Duration

REFERENCES

Hershfield, D. M., 1961: Rainfall frequency atlas of the United States for durations from 30 minutes to 24 hours and return periods from 1 to 100 years, *Technical Paper No. 40*, U. S. Weather Bureau, Washington, D.C.

Hosking, J.R.M. and J.R. Wallis, 1997: Regional Frequency Analysis, An approach based on L-moments, Cambridge University Press, NY, 224 pp.

Lin, B. and J.L. Vogel, A comparison of L-moments with method of moments, *Engineering Hydrology Symposium Proceedings*, ASCE, San Francisco, California, July 25-30, 1993.

Appendix

The Appendix contains briefing materials from the TP40/Ohio River Basin Study Interagency Meeting held 29 June 1999 at the Office of Hydrology, Silver Spring, Maryland.

Please Note: Acronym confusion - Lognormal (LNO) distribution should be Generalized normal (GNO).